

# Higher order sigma point filter: a new heuristic for nonlinear time series filtering

KSENIA PONOMAREVA, PARESH DATE

Department of Mathematical Sciences, Brunel University

---

## Abstract

In this paper we present some new results related to the higher order sigma point filter (HOSPof), introduced in [1] for filtering nonlinear multivariate time series. This paper makes two distinct contributions. Firstly, we propose a new algorithm to generate a discrete statistical distribution to match exactly a specified mean vector, a specified covariance matrix, the average of specified marginal skewness and the average of specified marginal kurtosis. Both the sigma points and the probability weights are given in closed-form and no numerical optimization is required. Combined with HOSPof, this random sigma point generation algorithm provides a new method for generating proposal density which propagates the information about higher order moments. A numerical example on nonlinear, multivariate time series involving real financial market data demonstrates the utility of this new algorithm. Secondly, we show that HOSPof achieves a higher order estimation accuracy as compared to UKF for smooth scalar nonlinearities. We believe that this new filter provides a new and powerful alternative heuristic to existing filtering algorithms and is useful especially in econometrics and in engineering applications.

**Keywords:** State estimation, sigma point filters, moment matching, nonlinear time series

---

## 1. Introduction

Consider the following state space form for a nonlinear time series:

$$\mathcal{X}(k+1) = \mathbf{f}(\mathcal{X}(k)) + \mathbf{Q}(\mathcal{X}(k))\mathbf{w}(k+1), \quad (1)$$

$$\mathcal{Y}(k) = \mathbf{h}(\mathcal{X}(k)) + \mathbf{v}(k), \quad (2)$$

where  $\mathcal{X}(k)$  and  $\mathcal{Y}(k)$  are the respective state vector and measurement vector at time  $t(k)$ ;  $\mathbf{f}$ ,  $\mathbf{h}$  are given vector-valued deterministic functions;  $\mathbf{Q}$  is a matrix valued deterministic function; and  $\mathbf{v}(k)$ ,  $\mathbf{w}(k)$  are vector-valued random variables. The time increment  $t(k) - t(k-1)$  is assumed constant for all  $k$ . The latent state estimation problem is the problem of constructing an estimate of the random vector  $\mathcal{X}(k)$ ,  $k \geq 1$ , based on the noisy time series data  $\mathcal{Y}(1), \mathcal{Y}(2), \dots, \mathcal{Y}(k)$ . In the special case when  $\mathbf{f}$ ,  $\mathbf{h}$  are affine in  $\mathcal{X}(k)$ ,  $\mathbf{Q}$  is an identity matrix and  $\mathbf{v}(k)$ ,  $\mathbf{w}(k)$  are Gaussian, the optimal recursive solution to the state estimation problem is given by linear Kalman filter, as first outlined in [2]. The optimal recursive solution to the state estimation problem in nonlinear systems is usually not available in closed form. The first current approach that addresses the nonlinear filtering problems is extended Kalman filter (EKF), where equation (1) or its continuous time analogue is locally linearized resulting in a linear state space system. A Kalman filter is

then employed to obtain the conditional state density of  $\mathcal{X}(k)$ . Standard textbooks such as [3] carry an extensive discussion of its theoretical underpinnings and implementation; also see [4]-[5]. The second method is unscented Kalman filter (UKF), where a set of particles - or *sigma points* - and weights are used to evaluate the terms in closed-form expressions for updating the state estimate. Several applications of UKF in communication, tracking and navigation are discussed in [6]-[7], among others. The *ensemble filter* (EF) used in climatology is closely related to UKF; see [8] and references therein. An algorithm which combines some of the desirable properties of both UKF and EF has been proposed in [9]. In [10], approximate methods are developed to deal with the multiplicative uncertainty in the observation equation under sigma point filtering framework. The third common method is sequential Monte Carlo filter or *particle filter* (PF). For this technique, the required conditional density function of  $\mathcal{X}(k)$  given measurement  $\mathcal{Y}(k)$  at time  $t(k)$  is represented by a set of random samples (or *particles*) and associated probability weights; see [11]-[12] and references therein for more details on PF. Particle filters need a specification of approximate posterior density, called the *proposal density*. This may itself be derived from EKF, UKF or the known state transition density. We will call the versions of particle filters as PF-EKF, PF-UKF and PF-T respectively.

The rest of this paper is organized as follows. Section 2 briefly reviews the traditional unscented Kalman filter. Section 3 introduces the algorithm for the unscented filter with higher order moment matching, which was first proposed by the authors in [1]. The traditional particle filter is discussed in section 4.1. Sections 4.2 and 6 represent the main contribution of this paper. Section 4.2 introduces the new proposal distribution that uses and propagates the information about higher order moments. This section represents a major modification on the algorithm proposed in [1] since it allows random draws of sigma points. We are aware that matching of the average of higher order moments may not add value if the average is over a very large state dimension. However, the class of applications where the dimension is five or less is still very large; in fact, it is unusual to find time series models with more than four latent states in econometrics and finance. The proposed algorithm can outperform traditional filtering algorithms in latent state estimation of nonlinear time series models where the departure from conditional Gaussianity of prior distribution is quite significant and the state dimension is low enough to make matching of the average kurtosis and the average skewness useful. This is illustrated by an example in section 5, where the utility of our method is compared with PF-T, PF-EKF and PF-UKF in a multivariate case. The theoretical accuracy of the conditional mean and the conditional variance estimation using the new method for the univariate case is discussed in section 6. Section 7 summarizes the results of the paper.

## 2. Unscented Kalman Filter

Consider the system of equations (1)-(2) with nonlinear functions  $\mathbf{f}$  and  $\mathbf{h}$ . The unscented filtering algorithm can be briefly described as follows. Suppose that at time  $t(k)$ , the mean  $\hat{\mathcal{X}}(k|k)$  and the covariance  $\mathbf{P}_{xx}(k|k)$  are available for the system in equation (1). Then  $2n + 1$  symmetric *sigma points* are chosen in the following way:

$$\mathcal{X}^{(0)}(k|k) = \hat{\mathcal{X}}(k|k), \mathcal{X}^{(i)}(k|k) = \hat{\mathcal{X}}(k|k) \pm (\sqrt{(n + \kappa)\mathbf{P}_{xx}})_i, \quad (3)$$

where  $i = 1, 2, \dots, n$ ,  $\kappa$  is a scaling parameter and  $(\sqrt{\mathbf{P}_{xx}})_i$  is the  $i^{th}$  column of the matrix square root of  $\mathbf{P}_{xx}$ . The probability weights vector  $\mathbf{W}$ , where the  $i^{th}$  component  $\mathbf{W}_i$  is associated with the  $i^{th}$  sigma point  $\mathcal{X}^{(i)}(k|k)$ , is defined as:

$$\mathbf{W}_0 = \frac{\kappa}{n + \kappa}, \mathbf{W}_i = \frac{1}{2(n + \kappa)}, \quad i = 1, 2, \dots, 2n. \quad (4)$$

The following result can then be verified by a straightforward algebraic manipulation (see, e.g. [11]):

*Proposition 1*

Sigma points and corresponding probability weights defined in (3) and (4) match the mean  $\hat{\mathcal{X}}(k|k)$  and the covariance  $\mathbf{P}_{xx}(k|k)$  exactly.

We compute the predicted mean of  $\mathcal{X}(k+1|k)$  using

$$\mathcal{X}^{(i)}(k+1|k) = \mathbf{f}(\mathcal{X}^{(i)}(k|k)), \quad \hat{\mathcal{X}}(k+1|k) = \sum_{i=0}^{2N} \mathbf{W}_i \mathcal{X}^{(i)}(k+1|k), \quad (5)$$

where  $\mathbf{W}_i$  are defined in (4). Covariance matrices  $\mathbf{P}_{xy}(k+1|k)$  and  $\mathbf{P}_{yy}(k+1|k)$  are calculated as

$$\mathbf{P}_{xy}(k+1|k) = \sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}^{(i)}(k+1|k) - \hat{\mathcal{X}}(k+1|k)) \mathbf{v}^{(i)}(k)^T, \quad \mathbf{P}_{yy}(k+1|k) = \sum_{i=0}^{2N} \mathbf{W}_i \mathbf{v}^{(i)}(k) \mathbf{v}^{(i)}(k)^T,$$

where  $\mathbf{v}^{(i)}(k) = \mathcal{Y}^{(i)}(k+1) - \hat{\mathcal{Y}}(k+1)$ ,  $\mathcal{Y}^{(i)}(k+1) = \mathbf{h}(\mathcal{X}^{(i)}(k+1|k))$  and  $\hat{\mathcal{Y}}(k+1) = \sum_{i=0}^{2N} \mathbf{W}_i \mathcal{Y}^{(i)}(k+1)$ .  $\mathbf{P}_{xx}(k+1|k)$  is computed similarly. Once the true measurement  $\mathcal{Y}_{k+1}$  becomes available, we can update the mean estimate in (5) as  $\hat{\mathcal{X}}(k+1|k+1) = \hat{\mathcal{X}}(k+1|k) + \mathbf{K}(k+1)(\mathcal{Y}_{k+1} - \hat{\mathcal{Y}}(k+1))$ , where  $\mathbf{K}(k+1) = \mathbf{P}_{xy}(k+1|k) \mathbf{P}_{yy}^{-1}(k+1|k)$ . More details on this algorithm can be found in [11]. UKF has been successfully used as an alternative to EKF; see [7]-[14] and references therein. Besides being used as a stand-alone filtering algorithm, it has also been used to produce a proposal distribution for PF, see [13].

Note that, while matching first and second moment accurately, UKF does not propagate any information about the 3<sup>rd</sup> and the 4<sup>th</sup> moments. This information may provide a better idea of the shape of the distribution and its departure from Gaussianity. Other suggested algorithms, which try to match higher moments, either require optimization or rely heavily on analytical solver, as in [15]. A new modification of generation of sigma points, first proposed in [1] is outlined in sections 3.1-3.2 for completeness.

### 3. A New Algorithm for Unscented Kalman Filtering

#### 3.1. Sigma point generation with higher order moment matching

In [16], a method was proposed to match the mean vector, the covariance matrix and the average marginal kurtosis of a multivariate distribution exactly, when the marginal densities are symmetric. In [1], this algorithm was extended to asymmetric distributions, as outlined below. We introduce additional parameters  $\alpha$  and  $\beta$  in order to capture the 3<sup>rd</sup> and the 4<sup>th</sup> moments of  $\mathcal{X}(k+1|k+1)$  using augmented UKF, which has been shown in [14] to give more accurate results compared to non-augmented UKF in the presence of significant noise terms. The augmentation method incorporates noise into the augmented random state vector and from here onwards we will assume  $\mathbf{f}$  and  $\mathbf{h}$  to be augmented functions. Suppose that we have  $\mathcal{X}$ , a random  $n$ -vector with mean  $\hat{\mathcal{X}}$  and covariance  $\mathbf{P}_{xx}$ , and noise  $\mathbf{w}$ , a random  $m$ -vector with zero mean and covariance  $\mathbf{Q}_{xx}$ , as in equation (1).  $n \neq m$  is allowed for generality. Matrices  $\mathbf{P} > 0$ ,  $\mathbf{Q} > 0$  are such that  $\mathbf{P}_{xx} = \mathbf{P}\mathbf{P}^T$  and  $\mathbf{Q}_{xx} = \mathbf{Q}\mathbf{Q}^T$ , where  $\mathbf{P}^T$  is transpose of  $\mathbf{P}$  and  $\mathbf{P}_i$  is the  $i^{\text{th}}$  column of matrix  $\mathbf{P}$ . We create  $2(n+m)+1$  sigma points as follows:

$$\begin{aligned} \mathcal{X}^{(0)} &= (\hat{\mathcal{X}} \quad \mathbf{0}_{m \times 1})^T = \bar{\mathcal{X}}, \quad \mathbf{W}_0 = 1 - \sum_{i=1}^{2N} \mathbf{W}_i \\ \mathcal{X}^{(i)} &= (\hat{\mathcal{X}} + \alpha \sqrt{N} \mathbf{P}_i \quad \mathbf{0}_{m \times 1})^T, \quad \mathbf{W}_i = \frac{1}{\alpha(\alpha + \beta)N}, \quad i = 1, 2, \dots, n, \\ \mathcal{X}^{(i)} &= (\hat{\mathcal{X}} - \beta \sqrt{N} \mathbf{P}_i \quad \mathbf{0}_{m \times 1})^T, \quad \mathbf{W}_i = \frac{1}{\beta(\alpha + \beta)N}, \quad i = n+1, \dots, 2n, \end{aligned}$$

$$\begin{aligned}\mathcal{X}^{(i)} &= (\hat{\mathcal{X}} \quad \sqrt{N}\mathbf{Q}_{i-2n})^\top, \quad \mathbf{W}_i = \left(\frac{1}{2N}\right), \quad i = 2n+1, \dots, 2n+m, \\ \mathcal{X}^{(i)} &= (\hat{\mathcal{X}} \quad -\sqrt{N}\mathbf{Q}_{i-2n})^\top, \quad \mathbf{W}_i = \left(\frac{1}{2N}\right), \quad i = 2n+m+1, \dots, 2n+2m,\end{aligned}\quad (6)$$

where  $N = n + m$ , the  $j^{th}$  element of a sigma point  $\mathcal{X}^{(i)}$  will be denoted as  $\mathcal{X}_j^{(i)}$ . Equations for  $\alpha$  and  $\beta$  are:

*Definition 1*

$$\alpha = \frac{1}{2}\phi_1 \pm \frac{1}{2}\sqrt{4\phi_2 - 3\phi_1^2}, \quad \beta = -\frac{1}{2}\phi_1 \pm \frac{1}{2}\sqrt{4\phi_2 - 3\phi_1^2}, \quad (7)$$

where the values of the same sign are taken,  $\phi_1 = \frac{\sum_{j=1}^n \omega_j}{\sqrt{N} \sum_{l=1}^n \sum_{k=1}^n \mathbf{P}_{lk}^3}$  and  $\phi_2 = \frac{\sum_{j=1}^n \psi_j}{N \sum_{l=1}^n \sum_{k=1}^n \mathbf{P}_{lk}^4}$ . Parameters  $\omega$  and  $\psi$  are vectors of the marginal 3<sup>rd</sup> and 4<sup>th</sup> central moments respectively, where  $\omega_j = \sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \bar{\mathcal{X}}_j)^3$  and  $\psi_j = \sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \bar{\mathcal{X}}_j)^4$ .  $\mathbf{P}_{ij}$  is the entry in the  $i^{th}$  row and the  $j^{th}$  column of matrix  $\mathbf{P}$ , so that  $\phi_1$  and  $\phi_2$  are known from data.

With the above choices of  $\alpha, \beta$ , we can match the mean, covariance, the average central third and fourth marginal moments as is stated in the next proposition:

*Proposition 2*

With  $\mathcal{X}^{(i)}, \mathbf{W}_i, \alpha, \beta$  chosen as in (6)-(7), the following properties hold:

$$\sum_{i=0}^{2N} \mathbf{W}_i = 1, \quad \sum_{i=0}^{2N} \mathbf{W}_i \mathcal{X}^{(i)} = (\hat{\mathcal{X}} \quad \mathbf{0}_{m \times 1})^\top, \quad (8)$$

$$\sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}^{(i)} - \bar{\mathcal{X}})(\mathcal{X}^{(i)} - \bar{\mathcal{X}})^\top = \begin{pmatrix} \mathbf{P}_{xx} & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & \mathbf{Q}_{xx} \end{pmatrix}, \quad (9)$$

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \bar{\mathcal{X}}_j)^3 = \frac{1}{n} \sum_{j=1}^n \omega_j, \quad (10)$$

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \bar{\mathcal{X}}_j)^4 = \frac{1}{n} \sum_{j=1}^n \psi_j. \quad (11)$$

**Proof:** Equations (8)-(9) follow by a straightforward algebraic manipulation. To prove equation (10), substituting expression for  $\mathcal{X}^{(i)}$  and  $\mathbf{W}_i$  from (6) into the left-hand side of equations (10) gives:

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \bar{\mathcal{X}}_j)^3 = \frac{\sqrt{N}}{n} \sum_{l=1}^n \sum_{k=1}^n \mathbf{P}_{lk}^3 \frac{(\alpha^2 - \beta^2)}{\alpha + \beta} = \frac{\sqrt{N}}{n} \sum_{l=1}^n \sum_{k=1}^n \mathbf{P}_{lk}^3 \phi_1 = \frac{1}{n} \sum_{j=1}^n \omega_j,$$

as required, where the last equality uses the definition of  $\phi_1$ . The equation (11) can similarly be proven by noting that

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{2N} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \bar{\mathcal{X}}_j)^4 = \frac{N}{n} \frac{(\alpha^3 + \beta^3)}{\alpha + \beta} \sum_{l=1}^n \sum_{k=1}^n \mathbf{P}_{lk}^4 = \frac{N}{n} (\alpha^2 - \alpha\beta + \beta^2) \sum_{l=1}^n \sum_{k=1}^n \mathbf{P}_{lk}^4 = \sum_{l=1}^n \sum_{k=1}^n \frac{N \mathbf{P}_{lk}^4 \phi_2}{n}$$

and using the definition of  $\phi_2$  provides the required result. ■

Note that  $\mathbf{W}_i \geq 0$  and  $\sum_{i=0}^{2N} \mathbf{W}_i = 1$  mean the set of probability weights and corresponding sigma points  $\{\mathbf{W}_i, \mathcal{X}^{(i)}\}$  forms a valid probability distribution. This is *not* always the case in unscented Kalman filter, since  $\kappa$  in (4) is not restricted to be positive. Provided  $\phi_2 \geq \frac{3}{4}\phi_1^2$ , (which is trivially true for symmetric distributions),  $\alpha$  and  $\beta$  allow us to capture and propagate the marginal skewness and marginal kurtosis. Note also that the unscented filter in section 2 employs the same weights  $\mathbf{W}_i$  for all sigma points  $\mathcal{X}^{(i)}$  for  $i > 0$ . In comparison, we have different expressions

for probability weights depending on  $i$  in (6). Higher order moment matching methods in the literature, such as that proposed in [15] and more recently in [17], rely on numerical optimization of nonlinear objective functions in several variables. In contrast, our algorithm presented here provides matching of average higher order moments in roughly the same amount of numerical effort as the UKF with augmented states, and further allows us to track changes in these higher order moments.

### 3.2. Filtering Algorithm using higher order moments

The filtering algorithm for higher order sigma point filter (HOSPof) can be described as follows.

1. Given  $\mathcal{X}^{(i)}(k|k)$ ,  $\mathbf{P}_{xx}(k|k)$  and the newly measured  $\mathcal{Y}_{k+1}$ , propagate sigma points using (5) to obtain  $\mathcal{X}^{(i)}(k+1|k) = \mathbf{f}(\mathcal{X}^{(i)})$ . Then using  $\hat{\mathcal{Y}}(k+1) = \sum_{i=0}^{2N} \mathbf{W}_i \mathcal{Y}^{(i)}(k+1)$ , with  $\mathbf{W}_i$  are the weights as defined in (6), the covariance matrices  $\mathbf{P}_{xy}(k+1|k)$  and  $\mathbf{P}_{yy}(k+1|k)$  can be calculated as described in section 2.
2. The propagated sigma points are updated as  $\mathcal{X}^{(i)}(k+1|k+1) = \mathcal{X}^{(i)}(k+1|k) + \mathbf{P}_{xy} \mathbf{P}_{yy}^{-1} (\mathcal{Y}_{k+1} - \mathbf{h}(\mathcal{X}^{(i)}(k+1|k)))$ , and the state estimate is updated using  $\hat{\mathcal{X}}(k+1|k+1) = \sum_{i=0}^{2N} \mathbf{W}_i \mathcal{X}^{(i)}(k+1|k+1)$ .
3. The updated sigma points  $\mathcal{X}^{(i)}(k+1|k+1)$  are used for the calculation of the  $\omega$  and  $\psi$ , the marginal  $3^{rd}$  and  $4^{th}$  central moments respectively at time  $t(k+1)$  via equations (10)-(11).
4. Definition 1 is used to obtain  $\phi_1$  and  $\phi_2$ . The updated values for  $\alpha$  and  $\beta$  are generated via equation (7).
5. The new set of sigma points and the corresponding weights at time  $t(k+1)$  can now be generated using (6).

## 4. A new proposal distribution

### 4.1. Traditional particle filter

One way to estimate the posterior distribution  $p(\mathcal{X}(k)|\mathcal{Y}(k))$  of the unobserved state  $\mathcal{X}(k)$  is by using particles drawn from it. Often posterior density is not known or it might not be easy to sample from it. We may instead choose to draw samples  $\mathcal{X}^{(i)}(k)$  from a known, easy-to-sample, *proposal* distribution  $q(\mathcal{X}(k)|\mathcal{Y}(k))$ . Given samples  $\mathcal{X}^{(i)}(k)$ , drawn from  $q(\mathcal{X}(k)|\mathcal{X}(k-1), \mathcal{Y}(k))$ , choosing the corresponding probability weights  $\mathbf{W}_i(k)$  such that

$$\mathbf{W}_i(k) \propto \frac{p(\mathcal{X}^{(i)}(k)|\mathcal{Y}(k))}{q(\mathcal{X}^{(i)}(k)|\mathcal{Y}(k))}, \quad (12)$$

ensures that  $\lim_{M \rightarrow \infty} \sum_{i=1}^M \mathbf{W}_i(k) \mathbf{h}(\mathcal{X}^{(i)}(k)) = \mathbb{E}(\mathbf{h}(\mathcal{X}^{(i)}(k)))$ , holds for any measurable function  $\mathbf{h}$  for which  $\mathbb{E}(\mathbf{h}(x))$  exists, where  $\mathbb{E}(\cdot)$  is expectation with respect to probability measure  $p(\mathcal{X}(k)|\mathcal{Y}(k))$ . The discrete distribution in PF is represented by a set of random particles and associated probability weights. The particles and weights are updated recursively as new measurements become available. In order to derive the recursive expression for updating the probability weights we assume the following factorization:

$$q(\mathcal{X}(k+1)|\mathcal{Y}(k+1)) = q(\mathcal{X}(k+1)|\mathcal{X}(k), \mathcal{Y}(k+1))q(\mathcal{X}(k)|\mathcal{Y}(k)). \quad (13)$$

Remembering that state dynamics is a Markov process and observations are conditionally independent given states  $p(\mathcal{X}(k+1)|\mathcal{Y}(k+1)) \propto p(\mathcal{Y}(k+1)|\mathcal{X}(k+1))p(\mathcal{X}(k+1)|\mathcal{X}(k))p(\mathcal{X}(k)|\mathcal{Y}(k))$ , when substituted together with (13) into (12) provides us with recursive estimates for the probability weights

$$\mathbf{W}_i(k+1) = \mathbf{W}_i(k) \frac{p(\mathbf{Y}(k+1)|\mathbf{X}^{(i)}(k+1))p(\mathbf{X}^{(i)}(k+1)|\mathbf{X}^{(i)}(k))}{q(\mathbf{X}^{(i)}(k+1)|\mathbf{X}^{(i)}(k), \mathbf{Y}(k+1))}. \quad (14)$$

Here  $p(\mathbf{Y}(k+1)|\mathbf{X}^{(i)}(k+1))$  and  $p(\mathbf{X}^{(i)}(k+1)|\mathbf{X}^{(i)}(k))$  denote the observation density and the state transition density respectively. Given  $\mathbf{X}^{(i)}(k)$ ,  $i = 1, 2, \dots, M$ ,  $\mathbf{X}^{(i)}(k+1)$  and  $\mathbf{W}_i(k+1)$  are obtained as:

- (a) Sample particles  $\mathbf{X}^{(i)}(k+1)$  from the proposal density  $q(\mathbf{X}(k+1)|\mathbf{X}(k), \mathbf{Y}(k+1))$ .
- (b) Compute the importance weights using (14).
- (c) Normalize the importance weights to obtain the new probability weights  $\widetilde{\mathbf{W}}_i(k+1) = \frac{\mathbf{W}_i(k+1)}{\sum_{i=1}^M \mathbf{W}_i(k+1)}$ .
- (d) Resample whenever a significant degeneracy is observed. Suitable criteria and algorithms for resampling could be found in [18].

Choosing the right proposal distribution is very important for a successful particle filter. Using the transition prior density  $p(\mathbf{X}^{(i)}(k+1)|\mathbf{X}^{(i)}(k))$  offers easy implementation due to (14), but does not incorporate most recent observations. Usually, a more efficient choice is Gaussian posterior density generated by the extended Kalman filter for the same system, i.e.  $q(\mathbf{X}(k)|\mathbf{Y}(k)) = N(\mathbf{X}(k|k), \mathbf{P}_{xx}(k|k))$ . Implementing posterior density obtained by UKF as the proposal distribution has been considered in [13], which seems to outperform EKF-based proposal in terms of the estimation accuracy.

We are interested in using information about higher order moments and not just mean and covariance (as in UKF or EKF-based proposals) for generating random particles. Continuing with  $\mathbf{W}_i$  and  $\mathbf{X}^{(i)}(k+1|k+1)$ , computed in HOSPoF, will limit us to a proposal with  $2N+1$  *deterministic* points or particles. Since PF-type filters work best for a large number of randomly generated particles, our aim is to find a non-Gaussian proposal which matches the given mean, the covariance matrix, the average marginal skewness and the average marginal kurtosis. Further, we aim to obtain the sigma points and the probability weights which achieve this moment matching in closed form, *e.g.* without needing an analytic solver or an optimizer. A method to achieve this, inspired by the particle filtering, is described next and is a further modification of HOSPoF outlined in section 3.

#### 4.2. A new moment-matching proposal distribution

We start by using the results from [9], where the authors defined a multivariate discrete distribution over  $2ns+1$  points and associated probability weights as follows:

*Definition 2*

$$\mathbf{X}^{(k)} = \hat{\mathbf{X}} \pm \frac{1}{\sqrt{2sp_i}} \mathbf{L}_j, \quad \mathbf{W}_k = \mathbf{p}_i, \quad (15)$$

$$\mathbf{X}^{(0)} = \hat{\mathbf{X}}, \quad \mathbf{W}_0 = \mathbf{p}_{s+1}, \quad (16)$$

where  $i = 1, 2, \dots, s$ ,  $j = 1, 2, \dots, n$ ,  $k = 1, 2, \dots, 2ns$ ,  $n$  is the dimension of the state vector  $\mathbf{X}$  as before,  $2n \sum_{i=1}^s \mathbf{p}_i + \mathbf{p}_{s+1} = 1$  and  $\mathbf{L}_j$  denotes the  $j^{th}$  column of a matrix  $\mathbf{L}$ .

By choosing the value of the parameter  $s$  we determine the size of the set of probability weights,  $p_i$ . If  $\mathbf{L}$  is a symmetric positive definite matrix such that  $\mathbf{L}\mathbf{L}^T = \mathbf{P}_{xx}$ , then sigma point filter, based on (15)-(16), matches the mean vector, the covariance matrix and zero third marginal moment exactly, see [9] for the proofs. Extending this method to non-symmetric distributions, we change (16) to introduce 2 new sigma points that would carry the information about average marginal third and fourth moments:

*Definition 3*

$$\mathbf{X}^{(k)} = \hat{\mathbf{X}} \pm \frac{1}{\sqrt{2s\mathbf{p}_i}} \mathbf{L}_j, \quad \mathbf{W}_k = \mathbf{p}_i, \\ \mathbf{X}^{(0)} = \hat{\mathbf{X}}, \quad \mathbf{W}_0 = \mathbf{p}_{s+1} \widetilde{\mathbf{W}}_0,$$

$$\begin{aligned}\mathcal{X}^{(2ns+1)} &= \hat{\mathcal{X}} + \frac{\tilde{\alpha}}{\sqrt{\mathbf{p}_{s+1}}} \mathbf{Z}, \quad \mathbf{W}_{2ns+1} = \mathbf{p}_{s+1} \tilde{\mathbf{W}}_1, \\ \mathcal{X}^{(2ns+2)} &= \hat{\mathcal{X}} - \frac{\tilde{\beta}}{\sqrt{\mathbf{p}_{s+1}}} \mathbf{Z}, \quad \mathbf{W}_{2ns+2} = \mathbf{p}_{s+1} \tilde{\mathbf{W}}_2.\end{aligned}\tag{17}$$

Here  $i, j, k$  are as in definition 2 and  $\mathbf{p}_i$  satisfy  $2n \sum_{i=1}^s \mathbf{p}_i + \mathbf{p}_{s+1} = 1$ ,  $\mathbf{Z}$  is an arbitrary non-zero deterministic vector. However  $\mathbf{Z}$  can also be randomly generated as long as it is non-zero and  $\mathbf{P}_{xx} - \mathbf{Z}\mathbf{Z}^T > 0$ . The matrix  $\mathbf{L}$  is such that  $\mathbf{L}\mathbf{L}^T + \mathbf{Z}\mathbf{Z}^T = \mathbf{P}_{xx}$  and  $\tilde{\mathbf{W}}_i$  satisfy  $\sum_{i=0}^2 \tilde{\mathbf{W}}_i = 1$  and are given by

$$\tilde{\mathbf{W}}_1 = \frac{1}{\tilde{\alpha}(\tilde{\alpha} + \tilde{\beta})}, \tilde{\mathbf{W}}_2 = \frac{1}{\tilde{\beta}(\tilde{\alpha} + \tilde{\beta})}, \tilde{\mathbf{W}}_0 = 1 - \frac{1}{\tilde{\alpha}\tilde{\beta}},$$

where the parameterization of  $\tilde{\mathbf{W}}_i$  follows the structure used in (6).

We aim to match the average marginal third and fourth moments,  $\frac{1}{n} \sum_{j=1}^n \tilde{\omega}_j$  and  $\frac{1}{n} \sum_{j=1}^n \tilde{\psi}_j$  respectively, which are defined in the same way as in section 3. The closed-form expression for  $\tilde{\alpha}$  and  $\tilde{\beta}$  are:

*Definition 4*

$$\tilde{\alpha} = \frac{1}{2} \tilde{\phi}_1 \pm \frac{1}{2} \sqrt{4\tilde{\phi}_2 - 3\tilde{\phi}_1^2}, \quad \tilde{\beta} = -\frac{1}{2} \tilde{\phi}_1 \pm \frac{1}{2} \sqrt{4\tilde{\phi}_2 - 3\tilde{\phi}_1^2},\tag{18}$$

where we take values of the same sign and

$$\tilde{\phi}_1 = \frac{\sum_{j=1}^n \tilde{\omega}_j \sqrt{\mathbf{p}_{s+1}}}{\sum_{k=1}^n \mathbf{Z}_k^3}, \quad \tilde{\phi}_2 = \mathbf{p}_{s+1} \frac{\sum_{j=1}^n \tilde{\psi}_j - \frac{1}{2s^2} \sum_{l=1}^n \sum_{k=1}^n \mathbf{L}_{lk}^4 (\sum_{i=1}^s \frac{1}{\mathbf{p}_i})}{\sum_{k=1}^n \mathbf{Z}_k^4}.$$

$2ns + 3$  randomly generated sigma points and corresponding probability weights defined in (17) match exactly the given mean, the covariance matrix, the average marginal third and fourth moments, *irrespective* of the precise choice of  $\mathbf{p}_i$ , as is stated in the next proposition:

*Proposition 3*

With  $\mathcal{X}^{(i)}$ ,  $\mathbf{W}_i$ ,  $\tilde{\alpha}$ ,  $\tilde{\beta}$  chosen as in (17)-(18), the following properties hold:

$$\sum_{i=0}^{2ns+2} \mathbf{W}_i = 1, \quad \sum_{i=0}^{2ns+2} \mathbf{W}_i \mathcal{X}^{(i)} = \hat{\mathcal{X}},\tag{19}$$

$$\sum_{i=0}^{2ns+2} \mathbf{W}_i (\mathcal{X}^{(i)} - \hat{\mathcal{X}}) (\mathcal{X}^{(i)} - \hat{\mathcal{X}})^\top = \mathbf{P}_{xx},\tag{20}$$

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{2ns+2} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \hat{\mathcal{X}}_j)^3 = \frac{1}{n} \sum_{j=1}^n \tilde{\omega}_j, \quad \frac{1}{n} \sum_{j=1}^n \sum_{i=0}^{2ns+2} \mathbf{W}_i (\mathcal{X}_j^{(i)} - \hat{\mathcal{X}}_j)^4 = \frac{1}{n} \sum_{j=1}^n \tilde{\psi}_j.\tag{21}$$

**Proof:** Equations (19)-(20) can be verified by a straightforward algebraic manipulation. Substituting expressions for  $\mathcal{X}^{(i)}$  and  $\mathbf{W}_i$  from (17) into left-hand side of equations (21) we get

$$\tilde{\alpha} - \tilde{\beta} = \frac{\sum_{j=1}^n \tilde{\omega}_j \sqrt{\mathbf{p}_{s+1}}}{\sum_{k=1}^n \mathbf{Z}_k^3}, \quad \tilde{\alpha}^2 - \tilde{\alpha}\tilde{\beta} + \tilde{\beta}^2 = \mathbf{p}_{s+1} \frac{\sum_{j=1}^n \tilde{\psi}_j - \frac{1}{2s^2} \sum_{l=1}^n \sum_{k=1}^n \mathbf{L}_{lk}^4 (\sum_{i=1}^s \frac{1}{\mathbf{p}_i})}{\sum_{k=1}^n \mathbf{Z}_k^4}.$$

On using definitions of  $\tilde{\alpha}$  and  $\tilde{\beta}$  from (18), we get the required result. ■

By imposing a constraint  $\tilde{\phi}_2 \geq \frac{3}{4} \tilde{\phi}_1^2$ , we obtain an upper limit for  $\sum_{i=1}^s \frac{1}{\mathbf{p}_i}$ , namely  $\zeta$ . The following proposition gives an expression for  $\zeta$ .

*Proposition 4*

$$\zeta = 2s^2 \frac{\sum_{j=1}^n \tilde{\psi}_j - \frac{3}{4} \sum_{k=1}^n \mathbf{Z}_k^4 (\frac{\sum_{j=1}^n \tilde{\omega}_j}{\sum_{k=1}^n \mathbf{Z}_k^3})^2}{\sum_{l=1}^n \sum_{k=1}^n \mathbf{L}_{lk}^4}.\tag{22}$$

**Proof:** Easy to verify by a straightforward algebraic manipulation. ■

Once  $\zeta$  is determined, the algorithm for generating  $2ns + 3$  sigma points, matching the given mean, the covariance matrix and the average marginal third and the average marginal fourth moments exactly, is provided below.

1. Choose an arbitrary vector  $\mathbf{Z}$  and compute  $\zeta$  using (22) and the given moments. Then generate  $\mathbf{p}_i > 0$  for  $i=1,2,\dots, s$  such that  $\sum_{i=1}^s \frac{1}{\mathbf{p}_i} < \zeta$ ,  $\mathbf{p}_{s+1} = 1 - 2n \sum_{i=1}^s \mathbf{p}_i$ . Note that  $\mathbf{p}_i$  can be generated using a random number generator, and  $s$  is an arbitrary positive integer. Hence the number of sigma points is independent of the state dimension. Note also that the distribution of  $\mathbf{p}_i$  is of no consequence since the moment-matching is independent of how  $\mathbf{p}_i$  are generated.
2. Calculate  $\tilde{\phi}_1$  and  $\tilde{\phi}_2$  for the chosen vector  $\mathbf{Z}$  using definition 3.
3. Closed form solution for  $\tilde{\alpha}$  and  $\tilde{\beta}$  are given by (18).
4. Construct the  $2ns + 3$  support points using (17).

The above sigma point generation algorithm can be used in the first step of the filtering algorithm in the section 4.1. However, it is worth outlining the differences between the traditional particle filter and the approach proposed here. First, while re-sampling is possible after moment matching scenario generation, the moments will no longer be exactly matched after re-sampling. This limitation is not specific to our method and is shared by all particle filters which use moment matching to generate a proposal. Secondly, the generated samples are not i.i.d but are ‘almost independent’, with some dependence introduced through the constraints on  $\mathbf{p}_i$ . For a large  $s$ , this dependence is minimal, and it may be preferable than generating a fixed number of points using a deterministic generator, as in the case of UKF. This intuition is supported by the numerical experiments in the next section. We denote this new method of using moment-matching proposal as PF-HOSPof.

## 5. Numerical Example - Multi-factor CIR model

For the empirical, ‘real-life’ example, we consider discretisation of a two-factor Cox-Ingersoll-Ross (CIR) model with a nonlinear measurement equation. This is a multivariable extension of the model first proposed in [19]; see [20] for more details on the use of this model in a filtering context. The state evolution is  $\mathcal{X}_j(k+1) = \kappa_j \epsilon_j \theta_j + (1 - \kappa_j \epsilon_j) \mathcal{X}_j(k) + \mathbf{Q}_j(k+1) \mathbf{w}_j(k+1)$ , for  $j = 1, 2$ , where  $\mathbf{w}_j(k)$  are zero mean, unit variance and uncorrelated Gaussian random variables. The standard deviation  $\mathbf{Q}$  is given by  $\mathbf{Q}_j(k+1) = \sigma_j \sqrt{\epsilon_j (\frac{1}{2} \theta_j \kappa_j \epsilon_j + (1 - \kappa_j \epsilon_j) \mathcal{X}_j(k))}$ , where  $\kappa_j$ ,  $\sigma_j$  and  $\theta_j$  are constants and  $\epsilon_j = \frac{(1 - e^{(-\kappa_j \Delta)})}{\kappa_j}$ , with  $\Delta = t(k+1) - t(k)$ . The observable variables are exponential in the latent states and are given by  $\mathcal{Y}_i(k) = \Pi_{j=1}^2 \left( \mathbf{A}_{i,j} \exp(-\sum_{j=1}^2 (\mathbf{B}_{i,j} \mathcal{X}_j(k))) \right) + \mathbf{z}_i(k)$ , where  $\mathbf{A}_{i,j} = \left( \frac{2\gamma_j \exp((\kappa_j + \gamma_j + \lambda_j)T_i/2)}{2\gamma_j + (\kappa_j + \lambda_j + \gamma_j)(\exp(T_i \gamma_j) - 1)} \right)^{\frac{2\kappa_j \theta_j}{\sigma_j^2}}$ ,  $\mathbf{B}_{i,j} = \frac{2(\exp(T_i \gamma_j) - 1)}{2\gamma_j + (\kappa_j + \lambda_j + \gamma_j)(\exp(T_i \gamma_j) - 1)}$  and  $\gamma_j = \sqrt{(\kappa_j + \lambda_j)^2 + 2\sigma_j^2}$ . The term  $\mathbf{z}_i(k)$  is observational noise with zero mean and is assumed to have a constant variance  $h^2$  for each  $i$  and  $\lambda_i$  are constants. In practice,  $T_i$  represents time to maturity and  $\mathcal{Y}_i(k)$  represents the price of a zero coupon bond with maturity  $T_i + t(k)$ , at time  $t(k)$ . Here, we use three maturities,  $T_1 = 1$ ,  $T_2 = 2$  and  $T_3 = 4$ . We use weekly data from February 2001 to July 2005 for UK government bonds. Here 180 observations were used for calibration and 52 were used for out-of-sample validation. A 2-factor model was calibrated using EKF and the quasi-maximum likelihood method. Table 1 presents the parameter values obtained as a result of calibration.



**Table 1. Parameter values**

$\theta_1$	$\theta_2$	$\sigma_1$	$\sigma_2$	$\kappa_1$	$\kappa_2$	$\lambda_1$	$\lambda_2$	$h$
0.0254	0.0175	0.0710	0.1870	0.0978	0.8035	-0.0350	-0.0490	0.001

After calibration, we use the sigma point generation method described in section 3 to generate sigma points at each  $t(k)$ , with initial values for mean  $\theta_j$  and diagonal elements of covariance as  $\frac{\theta_j \sigma_j^2}{2\kappa_j}$ . Eleven sigma points are generated at each  $t(k)$ . We construct  $\hat{\mathcal{X}}_j(k|k)$ ,  $j = 1, 2$  and  $\mathbf{P}_{xx}(k|k)$ , which are employed as proposal density for particle filter in order to obtain predictions of  $\mathcal{Y}_i(k)$ ,  $i = 1, 2, 3$ . As a benchmark for comparison, we use the predictions made using the generic particle filter with approximate transition density with 1000 particles and traditional augmented UKF combined with particle filter as described in section 2 and 4. In addition we run PF-HOSPof with 203 sigma points. Note that due to our unequal probability weights we can approximate the ‘shape’ of the density function by using a smaller number of points. Errors in this example are computed as

$$MRAE_{(i)} = \frac{1}{F} \sum_{j=1}^F \frac{|\mathcal{Y}_i(j) - \hat{\mathcal{Y}}_i(j)|}{\mathcal{Y}_i(j)}, \quad RMSE_{(i)} = \sqrt{\frac{1}{F} \sum_{j=1}^F (\mathcal{Y}_i(j) - \hat{\mathcal{Y}}_i(j))^2},$$

where the subscript  $i$  denotes  $i^{th}$  time to maturity. The out-of-sample results for UKF and HOSPof as stand-alone filters (without a particle filtering procedure) are included in this paper for completeness; see table 2 below. Tables 3-4 list the errors computed for one step ahead prediction, where results are averaged over 100 runs. These are a particle filter with approximate transition density as proposal (PF-T), a particle filter with Gaussian proposal generated by unscented Kalman filter (PF-UKF), by the extended Kalman filter (PF-EKF) and finally our new filter with a possibly non-Gaussian proposal using the new moment matching heuristic proposed here (PF-HOSPof).

From the tables, it can be seen that the PF-HOSPof outperforms PF-T, PF-EKF and PF-UKF for all the three times to maturity for out-of-sample data. The results on in-sample prediction are consistent with those on out-of-sample prediction and are omitted for brevity. The average improvement achieved with PF-HOSPof is about 26% for all the maturities, as compared to PF-UKF. This improvement is obtained with very little extra computational effort, viz. computing the marginal  $3^{rd}$  and  $4^{th}$  moments and  $\alpha, \beta$  using the closed-form expression in (7). All the numerical experiments were performed using Matlab 7.2 on a desktop with a dual core Pentium IV processor, 2.40GHz and 3.24Gb RAM. We have also included in table 5 a summary of average (over 100 runs) computational times taken for one run of PF-UKF and PF-HOSPof, and the average improvement for both types of errors achieved by PF-HOSPof over PF-UKF.

**Table 2. Average errors 1 step ahead out-of-sample prediction - standalone filters (without particle filtering)**

$\tau_k$	UKF-RMSE	HOSPof-RMSE	UKF-MRAE	HOSPof-MRAE
1Y	0.00087	0.00083	0.00075	0.00066
2Y	0.00147	0.00140	0.00144	0.00125
4Y	0.00283	0.00266	0.00300	0.00268

**Table 3. Average root mean squared error in 1 step ahead out-of-sample prediction**

$\tau_k$	PF-T	PF-EKF	PF-UKF	PF-HOSPof
1Y	0.006950	0.000902	0.000162	0.000126
2Y	0.010574	0.001218	0.000220	0.000142
4Y	0.016215	0.002224	0.000257	0.000209

**Table 4. Average relative absolute error of 1-step ahead out-of-sample prediction**

$\tau_k$	PF-T	PF-EKF	PF-UKF	PF-HOSPof
1Y	0.006789	0.000863	0.000141	0.000105
2Y	0.010422	0.001158	0.000191	0.000126
4Y	0.017335	0.002237	0.000244	0.000195

**Table 5. Comparison of accuracy and computational time for PF-UKF and PF-HOSPof**

Time taken on average for one run, s		Time improvement of PF-HOSPof over PF-UKF	Accuracy improvement of PF-HOSPof over PF-UKF	
PF-UKF	PF-HOSPof		AvRMSE	AvMRAE
29.1	5.7	80.4%	25.5%	26.5%

## 6. Theoretical Accuracy of higher order moment matching filter

In this section, we consider the theoretical accuracy of the sigma point generation algorithm used in HOSPof in the scalar case with smooth nonlinearities, in terms of estimating the conditional mean and the conditional variance.

### 6.1. Mean Estimation

We consider univariate case and as in section 3 we start with a random variable  $\mathcal{X}$  with mean  $\hat{\mathcal{X}}$  and variance  $\mathbf{P}_{xx}$ . Suppose that the second random variable  $\mathcal{Z}$  is related to  $\mathcal{X}$  through the nonlinear transformation  $\mathcal{Z} = \mathbf{f}(\mathcal{X})$ . Then the expected value of  $\mathcal{Z}$  is:

$$\hat{\mathcal{Z}} = \mathbf{f}(\hat{\mathcal{X}}) + \frac{1}{2!}\mathbf{f}''\mathbf{P}_{xx} + \frac{1}{3!}\mathbf{f}'''\mathbb{E}(\mathcal{X} - \hat{\mathcal{X}})^3 + \frac{1}{4!}\mathbf{f}''''\mathbb{E}(\mathcal{X} - \hat{\mathcal{X}})^4 + \dots \quad (23)$$

Now using the new sigma points,  $\mathcal{X}^{(i)}$ , and the corresponding probability weights,  $\mathbf{W}_i$  introduced in section 3, we analyze the accuracy in estimating the mean  $\hat{\mathcal{Z}} = \sum_{i=0}^{2N} \mathbf{W}_i \mathbf{f}(\mathcal{X}^{(i)})$ .

*Proposition 5*

If the function  $\mathbf{f}$  is at least five times differentiable, then

$$\hat{\mathcal{Z}} = \mathbf{f}(\hat{\mathcal{X}}) + \frac{1}{2!}\mathbf{f}''\mathbf{P}_{xx} + \frac{1}{3!}\mathbf{f}'''\omega + \frac{1}{4!}\mathbf{f}''''\psi + \dots, \quad (24)$$

where, as per section 3,  $\omega$  and  $\psi$  are the marginal 3rd and 4th central moments respectively, the derivatives  $\mathbf{f}''$  etc are computed at  $\mathcal{X} = \hat{\mathcal{X}}$ .

**Proof:** Expanding the expression for  $\hat{\mathcal{Z}}$  for the minimum possible number of sigma points ( $= 2N + 1 = 3$ ), we get  $\hat{\mathcal{Z}} = \sum_{i=0}^{2N} \mathbf{W}_i \mathbf{f}(\mathcal{X}^{(i)}) = \mathbf{W}_0 \mathbf{f}(\hat{\mathcal{X}}) + \mathbf{W}_1 \mathbf{f}(\hat{\mathcal{X}} + \alpha\sqrt{N\mathbf{P}_{xx}}) + \mathbf{W}_2 \mathbf{f}(\hat{\mathcal{X}} - \beta\sqrt{N\mathbf{P}_{xx}})$ . Further expanding  $\mathbf{f}$  around the mean  $\hat{\mathcal{X}}$ ,

$$\begin{aligned} \hat{\mathcal{Z}} = & \mathbf{W}_0 \mathbf{f}(\hat{\mathcal{X}}) + \mathbf{W}_1 \left( \mathbf{f}(\hat{\mathcal{X}}) + \mathbf{f}'\alpha\sqrt{\mathbf{P}_{xx}} + \frac{1}{2!}\mathbf{f}''(\alpha\sqrt{\mathbf{P}_{xx}})^2 + \frac{1}{3!}\mathbf{f}'''(\alpha\sqrt{\mathbf{P}_{xx}})^3 + \frac{1}{4!}\mathbf{f}''''(\alpha\sqrt{\mathbf{P}_{xx}})^4 + \dots \right) \\ & + \mathbf{W}_2 \left( \mathbf{f}(\hat{\mathcal{X}}) - \mathbf{f}'\beta\sqrt{\mathbf{P}_{xx}} + \frac{1}{2!}\mathbf{f}''(\beta\sqrt{\mathbf{P}_{xx}})^2 - \frac{1}{3!}\mathbf{f}'''(\beta\sqrt{\mathbf{P}_{xx}})^3 + \frac{1}{4!}\mathbf{f}''''(\beta\sqrt{\mathbf{P}_{xx}})^4 + \dots \right). \end{aligned}$$

Collecting similar powers of  $\mathbf{P}_{xx}$  together gives  $\hat{\mathcal{Z}} = \mathbf{f}(\hat{\mathcal{X}}) + \mathbf{f}'\sqrt{\mathbf{P}_{xx}}[\mathbf{W}_1\alpha - \mathbf{W}_2\beta] +$

$$\frac{1}{2!}\mathbf{f}''\mathbf{P}_{xx}[\mathbf{W}_1\alpha^2 + \mathbf{W}_2\beta^2] + \frac{1}{3!}\mathbf{f}'''\sqrt{\mathbf{P}_{xx}}^3[\mathbf{W}_1\alpha^3 - \mathbf{W}_2\beta^3] + \frac{1}{4!}\mathbf{f}''''\mathbf{P}_{xx}^2[\mathbf{W}_1\alpha^4 + \mathbf{W}_2\beta^4],$$

and using (6) simplifies  $\hat{\mathcal{Z}}$  to  $\hat{\mathcal{Z}} = \mathbf{f}(\hat{\mathcal{X}}) + \frac{1}{2!}\mathbf{f}''\mathbf{P}_{xx} + \frac{1}{3!}\mathbf{f}'''\sqrt{\mathbf{P}_{xx}}^3[\alpha - \beta] + \frac{1}{4!}\mathbf{f}''''\mathbf{P}_{xx}^2[\alpha^2 - \alpha\beta + \beta^2] + \dots$ . Expression (7) helps us eliminate  $\alpha$  and  $\beta$  and obtain the required result. ■

Comparing (23) with (24), we conclude that the mean  $\hat{\mathcal{Z}}$  is calculated accurate up to 4<sup>th</sup> order using our new method for generating sigma points. This 4<sup>th</sup> order accuracy is preserved in the univariate augmented case as well. This is a great improvement compared to the 2<sup>nd</sup> order accuracy achieved in [6].

## 6.2. Variance Estimation

For accuracy in variance estimation for  $\mathbf{Z}$  we first consider:

$$\mathbf{P}_{zz} = (\mathbf{f}')^2 \mathbf{P}_{xx} - \frac{1}{4}(\mathbf{f}'')^2 \mathbf{P}_{xx}^2 + \mathbf{f}'\mathbf{f}''\mathbb{E}(\mathcal{X} - \hat{\mathcal{X}})^3 + \left(\frac{1}{3}\mathbf{f}'\mathbf{f}''' + \frac{1}{4}\mathbf{f}''^2\right)\mathbb{E}(\mathcal{X} - \hat{\mathcal{X}})^4 + \dots \quad (25)$$

Using our sigma points and probability weights as in section 3 we can write covariance  $\mathbf{P}_{zz}$  as

*Proposition 6*

$$\mathbf{P}_{zz} = (\mathbf{f}')^2 \mathbf{P}_{xx} - \frac{1}{4}(\mathbf{f}'')^2 \mathbf{P}_{xx}^2 + \mathbf{f}'\mathbf{f}''\boldsymbol{\omega} + \left(\frac{1}{4}(\mathbf{f}'')^2 + \frac{1}{3}\mathbf{f}'\mathbf{f}'''\right)\boldsymbol{\psi} + \dots \quad (26)$$

**Proof:** Start with  $\mathbf{P}_{zz} = \sum_{i=0}^{2N} \mathbf{W}_i(\mathbf{Z}^{(i)} - \hat{\mathbf{Z}})^2 = \mathbf{W}_0(\mathbf{f}(\hat{\mathcal{X}}))^2 + \mathbf{W}_1(\mathbf{f}(\hat{\mathcal{X}} + \alpha\sqrt{N\mathbf{P}_{xx}}))^2 + \mathbf{W}_2(\mathbf{f}(\hat{\mathcal{X}} - \beta\sqrt{N\mathbf{P}_{xx}}))^2 - \hat{\mathbf{Z}}^2$ . Applying (24) for  $\hat{\mathbf{Z}}$  and after simplifying we get

$$\begin{aligned} \mathbf{P}_{zz} &= (\mathbf{f}')^2 \mathbf{P}_{xx} [\mathbf{W}_1\alpha^2 + \mathbf{W}_2\beta^2] + \mathbf{f}'\mathbf{f}''\sqrt{\mathbf{P}_{xx}}^3 [\mathbf{W}_1\alpha^3 - \mathbf{W}_2\beta^3] - \mathbf{f}'\mathbf{f}''\sqrt{\mathbf{P}_{xx}}^3 [\mathbf{W}_1\alpha - \mathbf{W}_2\beta] \\ &+ \left(\frac{1}{4}(\mathbf{f}'')^2 + \frac{1}{3}\mathbf{f}'\mathbf{f}'''\right) \mathbf{P}_{xx}^2 [\mathbf{W}_1\alpha^4 + \mathbf{W}_2\beta^4] - \frac{1}{4}(\mathbf{f}'')^2 \mathbf{P}_{xx}^2 + \dots \end{aligned}$$

Getting rid off  $\alpha$  and  $\beta$  through (7) will yield the required result. ■

Comparing (25) with (26) we can conclude that our new sigma point filter estimates variance correct up to the 4<sup>th</sup> order for the univariate case. This is an improvement not only over accuracy to the 1<sup>st</sup> order of the traditional unscented transformation method in [6], but also over 2<sup>nd</sup> order accuracy in scaled unscented transformation in [16].

## 7. Conclusion

The contribution of this paper is twofold. Firstly, we have proposed a new filtering algorithm, in which the sigma points and the corresponding probability weights are modified at each step to match exactly the predicted values of the average marginal skewness and the average marginal kurtosis, besides matching the mean and covariance matrix. We have further extended this heuristic to a filter with a non-deterministic, possibly non-Gaussian moment matching proposal. Secondly, in the univariate case, we have shown that the HOSPoF predicts mean and variance up to fourth order of accuracy, which is a significant improvement on a similar result for UKF. A numerical example on real financial market data illustrates the utility of the proposed algorithm. We believe that HOSPoF and PF-HOSPoF would both be useful in engineering, econometrics and finance applications.

## References

- [1] P. DATE AND K. PONOMAREVA (2010) A new unscented Kalman filter with higher order moment-matching, *Proceedings of Mathematical Theory of Networks and Systems (MTNS 2010)*, Budapest.
- [2] R.E. KALMAN (1960) A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, **82**, 35–45.
- [3] B. ANDERSON AND J. MOORE (1979) Optimal filtering, Prentice-Hall.
- [4] A. H. JAZWINSKI (1970) Stochastic processes and filtering theory, Academic press.
- [5] H. W. SORENSON (1985) Kalman filtering: theory and application, IEEE press.
- [6] S. JULIER AND J. UHLMANN (2004) Unscented filtering and nonlinear estimation, *Proceedings of the IEEE*, **92**, 401–422.
- [7] J.L. CRASSIDIS AND F.L. MARKLEY (2003) Unscented filtering for spacecraft attitude estimation, *AIAA Journal on Guidance, Control and Dynamics*, **26**, 536–542.

- [8] G. EVENSEN (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geography*, **10**, 143–162.
- [9] P. DATE, L. JALEN AND R. MAMON (2008) A new algorithm for latent state estimation in non-linear time series models, *Applied Mathematics and Computation*, **203**, 224–232.
- [10] A. HERMOSO-CARAZO AND J. LINARES-PEREZ (2007) Different approaches for state filtering in nonlinear systems with uncertain observations, *Applied Mathematics and Computation*, **187**, 708–724.
- [11] M.S. ARULAMPALAM, S. MASKELL, N. GORDON AND T. CLAPP (2002) A tutorial on particle filters for online nonlinear /non-gaussian Bayesian tracking, *IEEE Transactions on Signal Processing*, **50**, 174–188.
- [12] F. DAUM (2005) Nonlinear filters: beyond the Kalman filter, *IEEE Aerospace and Electronic Systems Magazine*, **20**, 57–69.
- [13] R. MERWE, A. DOUCET, N. FREITAS AND E. WAN (2000) The unscented particle filter, *Technical Report CUED/F-INFENG/TR*, **380**.
- [14] Y. WU, D. HU, M. WU, AND X. HU (2005) Unscented Kalman filtering for additive noise case: augmented versus non-augmented, *IEEE Signal Processing Letters*, **12**, 357–360.
- [15] D. TENNE AND T. SINGH (2003) The higher order unscented filter, *Proceedings of American Control Conference*, **3**, 2441–2446.
- [16] S. JULIER (2002) The scaled unscented transformation, *Proceedings of the American Control Conference*, **6**, 555–559.
- [17] W. LEDERMANN, C. ALEXANDER AND D. LEDERMANN (2011) Random orthogonal matrix simulation, *Linear Algebra and its Applications*, **434**, 1444–1467.
- [18] A. DOUCET, N. DE FREITAS, AND N. GORDON (2001) Sequential Monte Carlo methods in practice, Springer.
- [19] J.C. COX, J.E. INGERSOLL AND S.A. ROSS (1985) A theory of the term structure of interest rates, *Econometrica*, **53**, 385–407.
- [20] A. GEYER AND S. PICHLER (1993) A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure, *Journal of Financial Research*, **22**, 107–130.